Estimation of movements in video images

The invention relates to a method of motion estimation in video image data, in which method, starting from a first and a second video image, parameter sets of two or more motion models are initially determined, and in which image objects are assigned to the motion models. The invention also relates to a device for performing the method, a device operating in accordance with the method for displaying video images, and a computer program product for motion estimation.

Advances in multimedia techniques have led to the development of a multitude of video formats and display standards. They are distinguished, inter alia, also by their image rate, i.e. the number of frames per unit of time. When a video sequence is to be displayed on a PC or TV display screen, it is necessary to adapt to the image rate of the display apparatus. Interfaces suitable for this purpose operate by means of conversion methods of a different complexity. The simplest method is to repeat or omit frames of the video sequence in the display, dependent on the desired image rate. However, when displaying video data thus treated, unwanted artifacts occur. Unwanted display errors occur dependent on the ratio of the image rates involved. The display appears to be jittery and irregular so that the motions displayed in the video sequence have an unnatural effect. More elaborate methods perform an interpolation between consecutive video images, in which an algorithm for motion estimation is used, which initially recognizes the displacements of individual pixels from one image to the other and generates image data therefrom which are temporally present between the images of the video sequence. The use of such methods in apparatus for home use requires the fundamental algorithms to supply a qualitatively high-value image rate conversion and require only a small number of computations because the digital signal processing electronics in apparatus for home use have a limited efficiency.

Motion estimation methods of the type described in the opening paragraph are not only suitable for image rate conversion but also for coding and compression in the transmission of video data, as well as for depth estimation in 3D image data processing, and for disparity estimations in stereo images.

Such a method is proposed in WO 99/16251. It is an efficient, object-oriented method of motion estimation in which two or more motion models are used so as to describe the displacement of image objects between a current and a previous video image. The motion models are determined by parameter sets from which displacement vectors can be computed. One of the motion models is needed to deal with those image parts which are static. The associated displacement vector is thus the zero vector. The parameter sets of the other motion models are determined by evaluating the match errors of the motion models in the description of the displacement of image objects between consecutive video images. For the interpolation, it is then necessary to segment the image data and assign appropriate motion models to the individual image objects. The result of the segmentation is separate objects, i.e. image parts which perform a similar or comparable displacement from the previous to the current video image.

The known motion estimation method is an efficient alternative to the otherwise conventional block-oriented method because the number of independently movable objects is small in normal video sequences and, consequently, only a correspondingly small number of motion models is to be processed. A small number of computations results therefrom, which renders the method universally usable, also for home use.

The fundamental object of the present invention is to further improve the known motion estimation method and simultaneously further reduce the complexity.

An important step in the motion estimation is the determination of the parameter sets for the motion models. In the known method, the parameter sets are combined to vectors. For each motion model, a parameter set is selected from a quantity of candidate vectors in accordance with a selection criterion. The selection criterion consists of the evaluation of a match error. This is computed as the sum of absolute differences of individual motion-compensated pixel intensities between the current and the previous video image, while a displacement vector in accordance with one of the candidate vectors is used for compensation. An essential problem is that it is not clear in advance which motion model is to be assigned to which image area and with which parameter set. The known method is performed in such a way that the above-mentioned selection criterion is initially used with all motion models for the same image areas. Then, without an assignment being fixed, the best fitting parameter sets are selected.

In accordance with the above-mentioned envisaged object, a further reduction of complexity in a motion estimation method of the type described in the opening paragraph is achieved in that only parts of the image area are taken into account when determining the

parameter sets. A problem then is that corresponding parts of the image area are to be selected in an appropriate manner so that displacements between the video images are as completely captured as possible. According to the invention, only those parts of the image area are thus taken into account for determining the parameter sets, in which the first video image is significantly distinguished from the second video image.

Such distinctions are a clear indication that there is a displacement at the corresponding locations. Image parts for determining the parameter sets can thereby be selected very easily without initially having to know more precise motion data. Moreover, it is avoided that stationary image parts are processed when determining the parameter sets for the motion models, for which stationary parts motion compensation is useless and therefore unnecessary. In fact, the parameters must be determined only in the non-stationary image parts.

Since the selection criterion is only used for parts of the image area, the number of required computations is greatly reduced so that the overall motion estimation is accelerated. Based on the fact that only some hardly moving objects are displayed in typical video sequences, it is sufficient under normal circumstances to limit oneself to a corresponding number of "interesting" points in the video image when determining the parameter sets.

The "interesting" image areas are suitably determined in that deviations between the video images are evaluated block by block, taking those blocks for determining the parameter sets into account in which the value of the deviation exceeds a predetermined threshold value. The image area is thus divided into individual blocks whose size should be dimensioned in such a way that the parameter sets can be determined by means of individual blocks. The deviations between the current and the previous video image may be determined, for example, by forming the absolute differences of the pixel intensities each time within the individual blocks. The result is a positive number so that it can be easily ascertained by comparison with a predetermined threshold value whether there is motion or no motion in the associated image area. When determining the parameter sets, the method according to the invention is limited to those blocks in which a given distinction between the two video images can be recognized on the basis of the pixel intensities.

This method has the additional advantage that the threshold value can be determined on the basis that the number of image areas taken into account for determining the parameter sets is limited to a predeterminable value. Since the overall method is to be performed in real time for the motion estimation, it should be ensured that the number of

computations remains below a fixed maximum value. It is thus possible to adjust the threshold value in the case of repeated use of the method according to the invention in such a way that the data processing time remains uncritical.

Practice has proved that it may be advantageous to take into account those parts of the image area for determining the parameter sets, in which motion was determined in previous video image data of a sequence of video images. A higher temporal consistency in the motion compensation is obtained in this way.

For performing the method according to the invention, a device for motion estimation in video image data is suitable, which device comprises a digital image memory in which a first and a second video image can be stored, and means for determining parameter sets of two or more motion models in accordance with a selection criterion. The device according to the invention comprises means for block-wise evaluation of the deviations between the current and the previous video image and for selection of those blocks for use of the selection criterion, in which the value of the deviation exceeds a predeterminable threshold value. Such devices may be used, for example, as components in television and video apparatuses. The digital image memory of the device according to the invention need not necessarily have a sufficient capacity for recording the first and the second video image simultaneously. The consecutive storage of the respective images is sufficient for the method according to the invention.

Devices for displaying video images such as, for example televisions, monitors etc., comprising a digital image memory in which video image data can be stored, and electronic means for processing the image data stored in the image memory and for displaying video images on a display device, the means for processing the image data comprising means for determining parameter sets of two or more motion models in accordance with a selection criterion, may advantageously benefit from the method according to the invention when the means for processing the image data further comprise means for block-wise evaluation of the deviations between the current and the previous video image and for selection of those blocks for use of the selection criterion, in which the value of the deviation exceeds a predeterminable threshold value. Conventional, digitally operating televisions and monitors may be operated in a simple manner in accordance with the method according to the invention, with an improvement of the quality of the image displayed. Devices in the sense mentioned above are, for example, the cathode ray tubes or dot matrix displays conventionally used in televisions and monitors. Other devices for visual display of digital image data are also feasible.

According to the invention, a computer program product is suitable for interpolation between pairs of video image data sets, which product comprises, as input, a first and a second video image and, starting therefrom, computes parameter sets of two or more motion models and supplies motion data describing the displacement of image objects from the previous to the current image, while the image data of the two video images are compared with each other and only those parts of the image area in which there are significant differences between the two video images are taken into account in the computation of the parameter sets. The computer program product may be made available on various data carriers such as diskettes, CD-ROMs or the like, but also for transfer via computer networks (for example, Internet).

These and other aspects of the invention are apparent from and will be elucidated with reference to the embodiments described hereinafter.

In the drawings:

Fig. 1 shows a selection of interesting image areas;

Fig. 2 is a block diagram of a motion estimation method according to the invention;

Fig. 3 is a block diagram of a device according to the invention, for displaying video images.

When determining the parameter sets for the motion models according to the invention, a selection criterion is applied to the selected image areas. The selection criterion consists of, for example, the evaluation of a match error ε. This is computed as the sum of absolute differences of individual motion-compensated pixel intensities between a current and a previous video image in the following manner:

$$\varepsilon(\vec{C}_o,n)=\sum_{\vec{x}\in I(n)} W_o(\vec{x})\cdot|F_s(\vec{x},n)-F_s(\vec{x}-\vec{C}_o(\vec{x},n),n-1)|$$

A summation is subsequently effected via image co-ordinates $\vec{x}=\begin{pmatrix}x\\y\end{pmatrix}$, comprised in a quantity $I(n)$ of selected image areas. The absolute differences between the pixel intensities in the current and previous video image are added to these image co-ordinates. $F_s(\vec{x},n)$ is the pixel intensity at the image co-ordinate $\vec{x}$ in a video image with a reduced raster. It has proved that, in determining the parameter, the use of a resolution-reduced (sub-sampled) image is entirely sufficient. This advantageously leads to a

considerable reduction of the number of computations. Due to the continuing index n, the number of the frame and hence the instant within the video sequence is indicated. $\vec{C}_o(\vec{x},n)$ indicates, for the image pair n, as a current image and n-1 as a previous video image, the displacement vector at the image co-ordinate $\vec{x}$ in accordance with the motion model with the index o. $W_o(\vec{x})$ represents a weighting factor with which it is taken into account which motion model o was assigned to the image co-ordinates $\vec{x}$ in earlier image data of the video sequence. A combination between the determination of the parameters and image segmentation can thereby be realized, which has advantages with a view to the temporal consistency of the motion estimation and the efficiency of the method.

Starting from four parameters, displacement vectors can be computed by means of the following motion model:

$$\vec{C}_o(\vec{x},n) = \begin{pmatrix} s_x(o,n) + x \cdot d_x(o,n) \\ s_y(o,n) + y \cdot d_y(o,n) \end{pmatrix}$$

This is a simple linear first-order model with which translations and scalings can be described. The model is determined by the parameter set

$$\vec{P}_o(n) = \left(s_x(o,n), d_x(o,n), s_y(o,n), d_y(o,n)\right)^T.$$

The parameter set is determined in such a way that the above-mentioned match error for the corresponding motion model o assumes a minimal value. In the motion estimation method according to the invention, at least two motion models are used every time, one of which always has the zero vector as a parameter set so that the stationary image areas are described by this motion model with the displacement vector $\vec{C}_0(\vec{x},n) = \vec{0}$.

The next step in the motion estimation according to the invention is the image segmentation, i.e. assigning image areas to the motion models. To this end, the overall image area is initially subdivided into blocks. In practice, quadratic blocks of 8x8 pixels have proved to be suitable. For all image co-ordinates within the blocks at the position $\vec{X}$, it then holds that $\vec{x} \in B(\vec{X})$. For each block, a match error is again computed on a motion model o:

$$\varepsilon_o(\vec{X},n) = \sum_{\vec{x} \in B(\vec{X})} | F_s(\vec{x} + (1-\alpha)\vec{C}_o(\vec{x},n),n) - F_s(\vec{x} - \alpha\vec{C}_o(\vec{x},n),n-1) |$$

The instant when the segmentation should be valid is determined by α. In the simplest case, that motion model o is assigned to the block $\vec{X}$ for which $\varepsilon_o(\vec{X},n)$ is minimal. The assignment is then filed in the segmentation mask $M(\vec{X},n)$.

According to the invention, the motion estimation method for determining the parameters of the motion models is limited to "interesting" image areas, given by the quantity $I(n)$. It is advantageous to fill the quantity $I(n)$ with those blocks that are in poor conformity with the corresponding blocks in a previous image. This may be effected, for example, in accordance with the following prescription:

$$I(n) = \left\{\vec{X} \mid \varepsilon_0(\vec{X}, n-1) \geq T\right\}$$

T is a predeterminable threshold value which fixes the extent of the deviation between two consecutive images as from which the parameters are determined in the relevant image area.

Fig. 1 is a video image 1 showing a motorcyclist 2 riding on a street 3. The motorcyclist moves from left to right in the section of the image. The background, thus also the street 3, is stationary. In the Figure, the selection of "interesting" image areas can be recognized and are shown as white blocks 4. A motion model describing the motion of the motorcyclist 2 is assigned to the white blocks 4. The image background is stationary and is assigned to another corresponding motion model.

Fig. 2 shows diagrammatically the motion estimation procedure in accordance with the invention. Starting from a previous video image 6, a current video image 7 and a threshold value 8, image areas that are interesting for determining the parameter sets are selected in a first step 9 of the method described above. All of these image areas are provided with weighting factors 10 for a plurality of motion models and subsequently further processed in a step 11 in which the parameters of the motion models are determined in accordance with a selection criterion. Starting from the completely determined motion models, the overall image area is then subdivided into blocks in a step 12, and the displacement vectors corresponding to the individual motion models are computed for each block. Subsequently, the image area is segmented in step 13, in which the blocks are assigned to the motion models. The assignments, which are included in the weighting 10 for the next image pair, are stored in a segmentation mask 14 which is then obtained.

Fig. 3 shows diagrammatically the structure of a digitally operating device which may be, for example, a television or a video monitor. The device receives a video signal 20 which is stored and prepared in a digital image processing unit 21. To this end, the image processing unit comprises an image memory 22, a processor 23 and a program memory 24. These elements may also be at least partly combined in a discrete component. The processor 23 runs through a program stored in the program memory 24, which program controls the image processing method according to the invention. A display unit 25 receives

image data 26 prepared by the image processing unit 21 and generates a signal 27 therefrom for driving a cathode ray tube 28 via which the video images are visually presented.